

# **INDUCTIVE PATTERN-BASED LAND USE/COVER CHANGE MODELS: A COMPARISON OF FOUR SOFTWARE PACKAGES**

## **Abstract**

Land use/cover change (LUCC), as an important factor in global change, is a topic that has recently received considerable attention in the prospective modeling domain. There are many approaches and software packages for modeling LUCC, many of them are empirical approaches based on past LUCC such as CLUE, DINAMICA, CA\_MARKOV and Land Change Modeler (both available in IDRISI). This study reviews the possibilities and the limits of these four modeling software packages. First, a revision of the methods and tools available for each model was performed, taking into account how the models carry out the different procedures involved in the modeling process: quantity of change estimate, change potential evaluation, spatial allocation of change, reproduction of temporal and spatial patterns, model evaluation and advanced modeling options. Additional considerations, such as flexibility, user friendliness, were also taken into account. Then, the four models were applied to a virtual case study to illustrate the previous descriptions with a typical LUCC scenario that consists of four processes of change (conversion of forest to two different types of crops, crop abandonment and urban sprawl) that follow different spatial patterns and are conditioned by different drivers. The outputs were compared to assess the quantity of change estimates, the change potential and the simulated prospective maps. Finally, we discussed some basic criteria to define a “good” model.

**Key words:** LUCC, modeling, GIS, virtual case study, prospective, simulation.

## **1. INTRODUCTION**

Land use/cover change (LUCC) is significant to a range of aspects of global environmental change and has, thus, received increasing attention from scientists and decision makers. Over the last two decades, a broad range of models of LUCC have been developed to assist in land management and to better understand, evaluate and project the future role of LUCC within the functioning of the earth system. Modeling, particularly if performed using a spatially explicit approach, is an important technique for projecting and exploring alternative future scenarios,

for conducting experiments that help understanding and for quantitatively describing key processes (Veldkamp and Lambin, 2001).

LUCC is closely associated with social, environmental, institutional, and economic processes, and LUCC models are, therefore, integrated and attempt to model the coupled human-environment system. LUCC models comprise a wide variety of methodological approaches, which can be categorized in different ways (Parker et al., 2003; Gaucherel and Houet, 2009). The models can be static or dynamic, spatial or non-spatial (i.e., exploring patterns of change vs. rates of change), deductive or inductive (i.e., with model parameters based on statistical correlations vs. explicit descriptions of the process), agent-based or pattern-based (i.e., emulation of individual decision makers vs. inference of underlying behavior from the observation of patterns in the LUCC). The models can use a large range of information (remotely sensed classified images, biophysical and socioeconomic variables, economic indicators and scenarios, census data, field survey, etc.), are often embedded in GIS and can eventually be used in combination (see, for example, Castella and Verburg, 2007 and Overmars et al., 2007). The models concerned by the present study are dynamic spatially explicit models and have standard procedure that is based on an inductive pattern-based approach. Important modeling approaches, such as agent based models (Parker et al., 2001), are therefore not included in the present study.

In this inductive pattern approach, LUCC is modeled empirically using past LUCC or land use/cover (LUC) spatial distribution to develop a mathematical model that estimates the change potential as a function of a set of explanatory spatial variables (Veldkamp and Lambin, 2001; Paegelow and Camacho Olmedo, 2008). This mathematical model can also be based on theoretical assumptions. In models based on past LUCC, usually two LUC maps from two different dates are compared to estimate the patterns and processes of change (types of transitions and rates of change) for model parameterization. . The analysis of past changes or LUC distributions with regard to spatial explanatory variables enables to assess the degree to which locations might likely change in the future (Kolb et al., 2013). This change potential is also referred to as propensity, probability, susceptibility and suitability in the literature. These maps by themselves can be considered as a first product generated in the modeling process. Further procedures, applied to create a prospective LUC map, involve techniques that are used

to allocate the amount of certain changes established through the projection of the amount of historical LUCC across space, and the procedures eventually reproduce the spatial patterns of changing landscapes. Finally, an assessment of the model performance is carried out and is often based on the spatial coincidence between a simulated map and an observed LUC map, which is generally obtained through remotely sensed image classification. Other methods include expert opinion, comparison of outputs generated with multiple models or multiple runs with the same model. The modeling process is shown in figure 1.

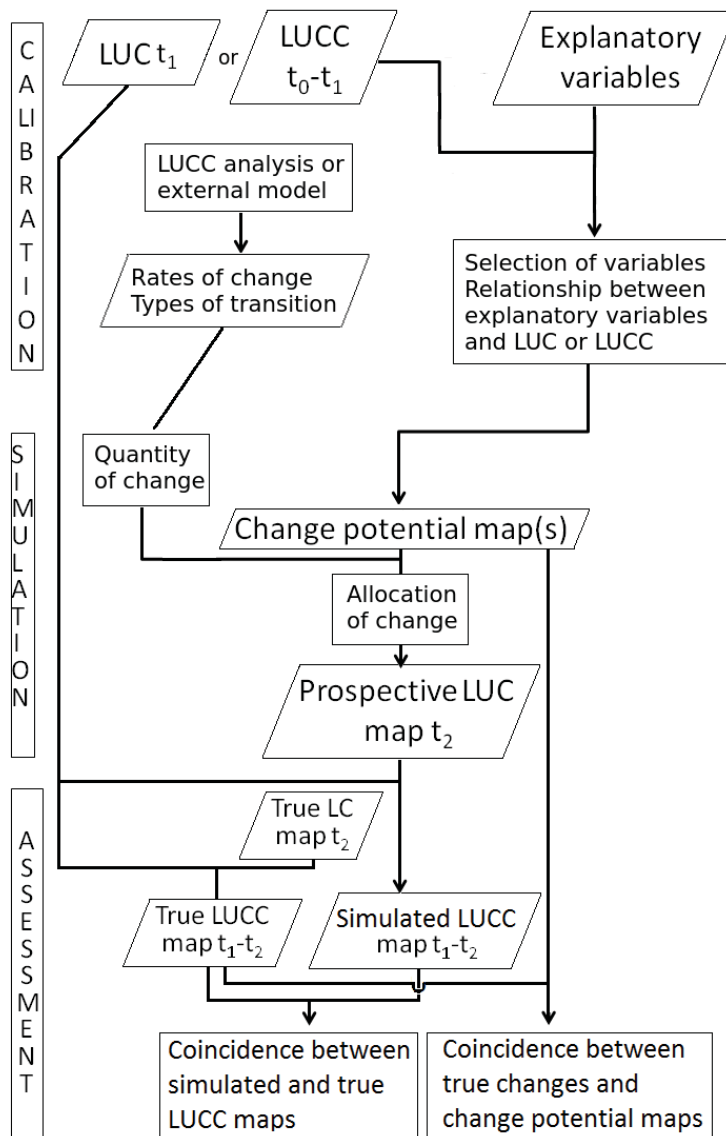


Figure 1 - Flowchart of the general procedure used in LUCC modeling. The rectangle shape indicates a process, the parallelogram inputs to and outputs from a process.

Different modeling software packages have certain functions that may be useful and appropriate depending on the available input data and the purpose of the LUCC modeling. Although some models have been widely used, and their performance has been evaluated more or less extensively (Agarwal et al., 2002; Eastman et al., 2005; Johnson, 2009; Pontius et al., 2008; Schaldach and Priess, 2008; Verburg et al., 2004b), there are very few studies aimed at understanding the benefits and limitations of modeling software programs by evaluating and comparing their tools and performance (Theobald and Hobbs, 1998; Pontius and Malanson, 2005; Castella and Verburg, 2007; Pocerwicz et al., 2008; Kim, 2010; Mas et al. 2011).

## **2. METHODS**

We compared and evaluated four frequently used models that are all based on an inductive pattern approach, but present important differences with respect to 1) the algorithm used to calibrate the model (from statistical to machine learning approaches), 2) the way they simulate the change (e.g. use or not of an automata cellular), 3) the methods used to assess the model performance and, 4) the flexibility to adapt the model to user's demands.

- IDRISI's CA\_MARKOV uses Markov chain matrices to determine the quantity of change along with suitability maps and cellular automata to spatially allocate these changes. A detailed application of this approach can be found in Paegelow and Camacho (2005). Examples of applications can be found in Houet et al. (2006), Poska et al. (2007), Shirley and Battaglia (2008), Kamusoko et al. (2009), Paegelow et al. (2008), Mobaied et al. (2011), Sang et al. (2011) and Adhikari and Southworth (2012).
- CLUE-S/Dyna-CLUE (Conversion of Land Use and Its Effects at Small regional extent) is based upon an analysis of location suitability using logistic regressions and simulates the competition and interactions between the different LUC types. More information on the development of this model can be found in Verburg et al. (2002) and Verburg and Overmars (2009). CLUE has been applied to a large variety of topics, including tropical deforestation (Verburg and Veldkamp, 2004, Wassenaar et al., 2007), biofuel crops (Hellmann and Verburg, 2009), farmland abandonment (Verburg and Overmars, 2009) and the effects of LUCC on carbon sequestration (Schulp et al., 2008).

- DINAMICA EGO uses transition probability maps that are based on the weight of evidence and genetic algorithm methods. These maps simulate landscape dynamics using both Markov chain matrices to determine the quantity of change and a cellular automata approach to reproduce spatial patterns. DINAMICA has been applied to a variety of studies, such as modeling urban growth (Almeida et al., 2003; Thapa and Murayama, 2011), tropical deforestation from local to basin-wide scales (Soares-Filho et al., 2002, 2006; 2013; Cuevas and Mas, 2008; Mas and Flamenco, 2011), fire regimes (Silvestrini et al., 2011; Soares-Filho et al., 2012), rent and opportunity cost of public policies or timber industry (Merry et al., 2009; Bowman et al., 2012; Giudice et al., 2012), the evaluation of the costs and benefits of reducing carbon emissions from deforestation and forest degradation (Nepstad et al., 2009), the assessment of the role of protected areas in reducing carbon emissions (Soares-Filho et al., 2010; Yanai et al., 2012) and the assessment of the co-benefits of REDD (Stickler et al., 2009, Nunes et al., 2012).
- Land Change Modeler (available in IDRISI or as an ARC-GIS extension) is a suite of tools with which the LUCC analysis and modeling can be combined with biodiversity and greenhouse gas emission assessments. The change modeling module is based on Markov chain matrices and transition susceptibility maps obtained by logistic regression or by training learning machines (Eastman, 2009; Johnson, 2009; Pineda-Jaimes, 2009). The Land Change Modeler was applied to identify trends in LUCC (Václavík and Rogan, 2009), tropical deforestation (Koi and Murayama, 2010), urban growth (Aguejdad and Houet, 2008), erosion under different conservation scenarios (Gaspari et al., 2009) and habitat modeling (Gontier et al., 2009).

Additionally, the R software environment for statistical computing and graphics was used to carry out the analysis of the results and elaboration of graphics of the virtual case study (R Development Core Team, 2012).

A review of the methods and tools offered by each model to perform LUCC simulations was carried out by taking into account the principal tasks involved in LUCC spatial modeling.

1) How does the model estimate the quantity of changes? 2) How does the model establish the relationship between the explanatory variables and the changes to assess the degree of change potential? 3) How does the model allocate the changes? 4) How does the model simulate spatial patterns? Finally, 5) Which tools are provided to evaluate the model? The first two tasks are related to the model calibration, the two following ones to the simulation and the last one to the model assessment (Figure 1). We also examined the possibility of developing more sophisticated models along with user friendliness and support materials.

In a subsequent step, each model was applied to a virtual case study, which is a data set created by the authors for this comparative study. This strategy made it possible to control the conditions and offered the possibility for elaborate, challenging situations to test specific modeling tasks. Therefore, the model assessment was not based upon the comparison between simulated and observed maps because such results largely depend on the specificities of the study area and change dynamics. In this simplified case study, the models were assessed using a range of criteria involving an analysis of the outputs (projection of the change magnitude, maps of the change potential, prospective LUC maps). The application to the virtual case was not intended to rank the models but rather to illustrate the different approaches used by the models to accomplish the modeling tasks. In the discussion we evaluate the flexibility/performance of each software package to accomplish the modeling tasks.

### **3. RESULTS**

#### **3.1. Model reviews**

##### **3.1.1. Modeling procedures**

In the following section, the methods and tools offered by each model to accomplish each one of the five principal tasks involved in LUCC spatial modeling are described. Subsequently, descriptions concerning advanced parameter settings and additional considerations are presented.

### 3.1.2. Quantity of change estimate

In CA\_MARKOV, DINAMICA and LCM, the changes are computed from a Markov matrix that is generally obtained through the comparison of the LUC maps from two dates. Markov chain projection provides the model with the estimated areas of each LUC category for future dates and the amount of change for each transition (“from-to” quantities).

The transition matrix for the period between dates  $t_0$  and  $t_1$  ( $t_1 = t_0 + T$ ) is obtained by overlaying the two LUC maps dated  $t_0$  and  $t_1$ . This matrix indicates the area (or number of pixels) for each transition and can be transformed into a Markov chain probability matrix for the entire period (hereafter the base transition matrix), which is the basis for projecting to a future date after one or several periods  $T$  (for instance date  $t_1 + T$ ). However, it is often desirable to use a time set different from the original period  $T$  for projecting into the future. For instance, the period between the two LUC maps used to calibrate the model is typically multiple years (10 years in our example) and the model runs using one year as the time step.

In DINAMICA, the base transition probability matrix is transformed into an annual matrix  $A$  by matricial calculation (Equation 1) to project the trends of change on an annual basis (Soares-Filho et al., 2002; Takada et al., 2010).

$$A = B^{1/t} = H \begin{pmatrix} (\lambda_1)^{1/t} & & 0 \\ & \ddots & \\ 0 & & (\lambda_n)^{1/t} \end{pmatrix} H^{-1} \quad (1)$$

where  $A$  is the annual matrix,  $t$  is the number of years,  $B$  is the original base transition matrix,  $H$  is the eigenvector of  $B$ , and  $\lambda_i$  is the  $i$ -th eigenvalue of  $B$ .

In IDRISI (CA\_MARKOV and LCM), the projections are also performed by creating a matrix to calculate the quantity of each LUC for a desired date. When the date being projected forward is a multiple of the calibration period, this transition probability matrix is calculated using a simple powering. For example, as in our virtual case study with a calibration period of

10 years (2000-2010), the base matrix squared would correspond to change probabilities for 2020, raised to the power of three to 2030 and so on. When the projected time period is not a multiple of the calibration period, then the power rule is used to generate three transition matrices that envelop the projection time period (the date to be interpolated will be between the first two dates). The three values at each cell in the transition probability matrix are then fed into a quadratic regression (thus, there is a separate regression for each cell), and this equation is then used to interpolate the unknown transition probability. A complete description of the algorithm can be found in the IDRISI Selva help system (command Markov).

CA\_MARKOV is the only model that takes into account the accuracy of the input LUC maps and uses this information to modify the transition matrix. When a proportional error greater than zero is specified, probability values that are equal to or greater than the values established for the probability of the permanence of that row in the matrix are reduced by the amount of the error (in percent). The remaining probability values (except null values) are evenly increased to force the probabilities to sum to one. Usually, this adjustment decreases the probability values of persistent pairs and increases the probability values of change pairs. However, in most cases, the mapping errors lead to an overestimation of the quantity of change, and the application of this corrective method will exacerbate the bias of the estimates of change probabilities.

In CLUE, the area of land cover categories (and not of each transition) is provided by the user for each simulated year. These areas can be obtained through a large variety of approaches (simple trend extrapolations, Markov projections, and economic models, among others) but should be evaluated using tools external to CLUE. In the absence of information about the quantity of change per transition, the indication of which LUC can be transformed into another is provided to the model through the concept of conversion elasticity and LUC type-specific transition sequences. The elasticity is related to the reversibility of LUCC. LUC with high capital investment or irreversible impact on the environment, such as in urban areas, are not easily converted into other LUC and, therefore, present a low elasticity. Other LUC are more easily converted and, therefore, have more elasticity (Verburg, 2010). A dimensionless value that represents the relative elasticity to conversion, which ranges from zero (easy conversion)



to one (irreversible change), is specified by the user based on expert knowledge or observed behavior in the recent past. The transition sequence is a set of rules that determine the possible LUC conversions.

### **3.1.3. Change potential evaluation**

Model calibration regarding change potential evaluation should assure that biophysical or socio-economical explanatory location characteristics create coherent change potential maps for LUCC simulations. Typical explanatory variables are the slope, the distance to roads and settlements, the land tenure and the soil types. The change potential of a given transition can be represented through two slightly different approaches: the suitability of a location for a given LUC resulting from the transition and the probability to present this transition.

CLUE and CA\_MARKOV use maps that express the suitability of a location for each of the LUC types. DINAMICA and LCM compute the probability of each transition taking place (in fact, when using neural networks or other machine learning tools, the values cannot be considered to be probabilities in a strict sense but are interpreted in the same manner, which is as values ranking the potential of change). The drawback of the suitability approach is that it does not consider the spatial configuration of past changes during the calibration period because it is based on the relationship of LUC to explanatory factors. For instance, a suitability map for a secondary forest will not indicate the more likely areas for the transition from forest to secondary forest (degradation) because this category can also be obtained through cropland abandonment (recovery), and these two processes are likely to occur at different places (Kolb et al., 2013). In this context CLUE uses the elasticity values to manage the amount of changes from one LUC to another, necessary to fulfill the established change rules. However, for a simulation over a long period of time with non-stationary change patterns, the suitability approach is likely to be more stable and give better results. Another advantage of this approach is that a model can be calibrated with only one available LUC map. In contrast, transition probability maps are derived from the relationships between areas that changed in the past and explanatory variables; therefore, these maps are more likely to be able to capture change processes and provide better spatial future LUCC estimations, which is

observed specifically for stationary change processes and short simulations (Kolb et al., 2013). Conversely, due to the reduced number of changes in relation to the entire study area, the statistical significance of the probability values can be affected, and this effect is less pronounced in the suitability approach.

Both types of change potential maps are elaborated by establishing a relationship between explanatory variables and LUC or transition types. In CA\_MARKOV, the suitability maps are generated by a multicriteria evaluation (MCE) for which the IDRISI framework provides multiple tools. CLUE uses logistic regression models that have to be run in separate programs for statistical analysis. It is also possible to use other models to elaborate the suitability maps, because these maps can be directly read by the model (Overmars et al., 2007). DINAMICA calculates a probability map using the weights of the evidence method (Bonham-Carter, 1994). The weights of the evidence can be tuned using expert knowledge or a genetic algorithm to improve the fit between the maps of the change probability and the LUCC map used for the model calibration (Soares et al., 2013). LCM offers three approaches to produce the probability map: logistic regression, multilayer perceptron (MLP) trained by backpropagation, which is a commonly used supervised neural network, and a similarity-weighted instance-based machine learning tool known as SimWeight (Sangermano et al., 2010).

These methods imply different assumptions and preprocessing of the explanatory variables. Logistic regression models and weights of evidence are based on the assumption of independence between explanatory variables. Such independence is often lacking, and DINAMICA and IDRISI have tools to evaluate the correlation between maps (Cramer's coefficient, Chi square, correlation and Kappa in IDRISI; Cramer's, contingency and joint information uncertainty in DINAMICA). The weight of evidence computing is based on categorical variables, and DINAMICA has a tool to preserve the data structure when converting continuous structures into a categorical map. Conversely, the logistic regression models, the multilayer perceptron and SimWeight use preferentially continuous variables because the conversion of a map of  $k$  categories into  $k-1$  categorical binary maps (dummy variables) increases the size of the model, affecting its performance. In IDRISI, categorical maps can be converted into continuous maps using the Evidence Likelihood transformation

based on the relative frequency of pixels belonging to the different categories within areas of change.

These methods offer different degrees of integration regarding expert knowledge and ability to fit the relationship between the explanatory variables and the transition potential or the suitability. Multicriteria evaluation (CA\_MARKOV) enables the consideration of expert knowledge, while logistic regression and especially machine learning tools (neural networks, genetic algorithms, SimWeight) do not. In DINAMICA, the relationships between explanatory variables and the weights of evidence can be displayed and eventually edited (Figure 2). It is, therefore, possible to tune the level of consideration of expert knowledge from an entirely statistical, data-driven approach (without modifying the statistically calculated values of the weights or modifying them through the genetic algorithm) to an expert knowledge-driven approach (important modification of the weights or complete edition by the expert). However, in all of these approaches, the use of expert knowledge is critical for the determination of the main processes, the selection of the potential driving variables and the evaluation of the outcomes of the change potential evaluation (Verburg et al., 2003). It is worth noting that even in machine learning, the user must specify the input maps based on hypotheses regarding possible determinants to avoid finding spurious correlations.

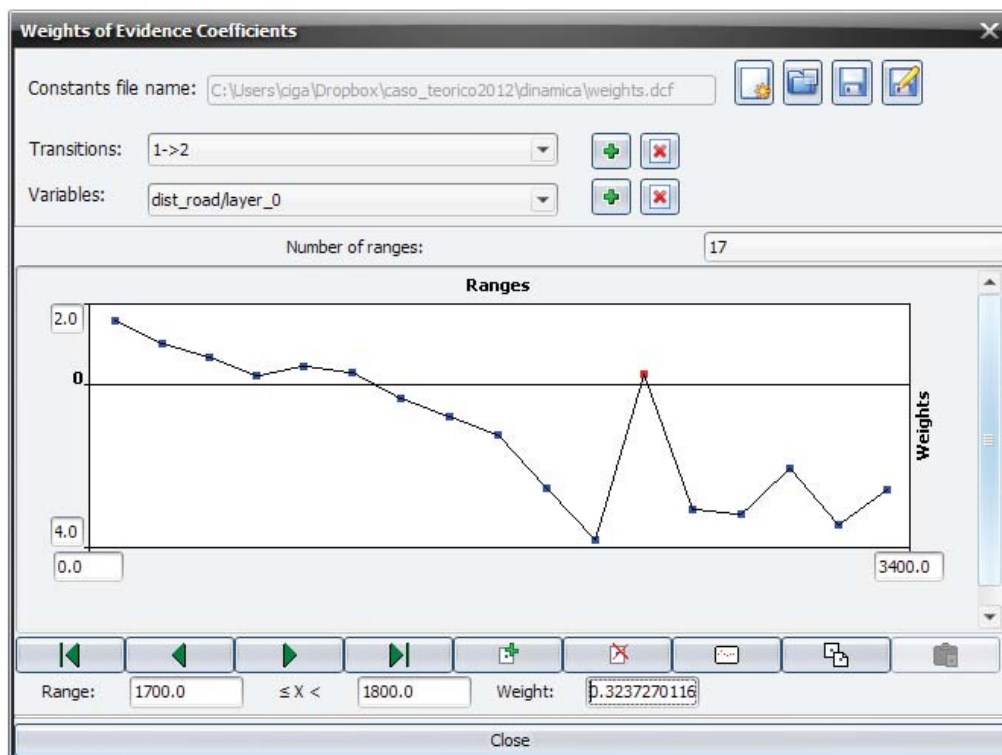


Figure 2 – Display and edition of weights of evidence (DINAMICA). This graph presents the weights of evidence of transition Forest to cropland 1 as a function of the distance from road. If the user would consider that the weight associated with the 12<sup>th</sup> range category is not an adequate representation of the effect of distance from road on the transition probability, he can modify its value.

The methods also differ in their flexibility to capture the relationships between the change potential and the explanatory variables. Logistic regression models will be unable to properly model a function that is not sigmoidal, but explanatory variables can be transformed to allow deviations from a sigmoid function. As the weights of evidence are calculated for each category, the weights can fit complex functions depending on the way the categories were defined. These two approaches are based on an additive effect of the explanatory variables. Conversely, machine learning approaches use model-free functions and can handle complex non-linear functions, taking into account the synergism or the inhibitory effects between variables. These approaches are, therefore, expected to better fit a function between change potential and explanatory variables. However, the capacity to model complex functions with a high goodness of fit is not necessarily an advantage in prospective modeling because it can lead to overfitting a model to the calibration period, such that the model will demonstrate a

poor performance when the past conditions (used to calibrate the model) and the conditions during the simulated period are not the same. For example, Mas et al. (2004) found that more complex neural networks allow for better fitting during the calibration period but fail in predicting change in the following period. Soares-Filho et al. (2009) observed the same overfitting effect using a genetic algorithm. DINAMICA enables the user to specify an envelope of maximum variation to overcome overfitting. Pérez-Vega et al. (2012) reported that the principal drawback when modeling deforestation was due to the differences between rates and spatial patterns of change during the calibration and the simulation periods. The multilayer perceptron used in LCM has different stopping criteria to avoid overfitting. In DINAMICA, the genetic algorithm constrains the new values of weights of evidence to an envelope around each statistically determined original value to avoid overfitting.

Finally, the approaches provide different levels of information to understand the effect of the explanatory variables on change and their interactions. From this point of view, logistic regression is perhaps the more useful approach. During regression model elaboration, a large number of methods and indices (stepwise selection of explanatory variables, Wald index, Akaike's Information Criterion, and analysis of variance, among others) can be used to select the variables with more predictive power based on their relative contribution to the model. However, the use of regression models to seek the best predictive model and identify factors influencing response variables (such as species or LUCC occurrence) may generate spurious results due to multicollinearity (MacNally, 2000 and 2002). The graphing of weights of evidence provided by DINAMICA also allows the visualization of the effects of the variables, but this graphing only takes into account one transition and one variable at a time, without considering the interaction between variables. The interpretation is very intuitive: a positive value indicates that the category favors the transition, while a negative value indicates an inhibition of the transition (low probability). For example, the values of the weights decrease with the distance from roads, and the difference in weight value between short and long distances indicates that the distance from roads has a strong effect on the transition probability (Figure 2). Machine learning approaches, such as neural networks, genetic algorithms or SimWeight, are often perceived as black boxes that establish a transfer function between inputs (the explanatory variables) and output (change potential) without any knowledge of

their internal working (Qiu and Jensen, 2004, Mas and Flores, 2008). This perception can be nuanced, such as in the case of SimWeight, which is simple to understand and provides the user with information on the relative importance of each explanatory variable (Sangermano et al., 2010). In order to train the IDRISI MLP, half of the training data are randomly selected for learning and half for validation. After the MLP has been trained, validation data are used to calculate a "skill measure" (computed as the accuracy of transition prediction *minus* the accuracy expected by chance). Different subsets of variables are maintained constant, are used to assess the contribution of individual variables, along with interaction effects.

Table 1 – Model features related to change potential evaluation. This table is based on the standard methods of each model, change potential maps obtained through other methods can be eventually incorporated into the models.

| Program   | Change potential map   | Analysis of drivers      | Expert Knowledge integration | Data driven | Expected goodness of fit (GF) / over fitting risk (OFR)                             |
|-----------|------------------------|--------------------------|------------------------------|-------------|---|
| CA_MARKOV | Suitability            | Multicriteria evaluation | Yes*                         | Yes*        | Depending on expert knowledge, satisfactory GF without OFR                          |
| CLUE      | Suitability            | Logistic regression      | No                           | Yes         | Poor GF if change potential is not a sigmoid function of the variable, OFR unlikely |
| DINAMICA  | Transition probability | Weight of evidence       | Yes*                         | Yes*        | High GF with OFR if change is statistically under-represented                       |
|           |                        | Genetic algorithm        | No                           | Yes         | High GF with OFR  |
| LCM       | Transition probability | Logistic regression      | No                           | Yes         | Poor GF if change potential is not a sigmoid function of the variable, OFR unlikely |
|           |                        | Multilayer perceptron    | No                           | Yes         | High GF with OFR  |
|           |                        | SimWeight                | No                           | Yes         | High GF with OFR  |

\* CA\_MARKOV and DINAMICA enable expert knowledge or data driven modelling depending on the way change potentials are created.

#### **3.1.4. Spatial allocation of change**

Change allocation is essentially a decision process that selects the pixels that actually change from one category to another from the change potential maps. Under the assumption that the pixels that will change are those that have the highest potential (or highest suitability for the “destination” LUC category), CA\_MARKOV and LCM rank the pixels and then select the top ranks according to the area required. Because there will commonly be competition for specific land parcels (the same location can be a candidate for various transitions), CA\_MARKOV and LCM use a multi-objective allocation procedure that iteratively reclassifies the ranked suitability maps to perform a first-stage allocation, checks for conflicts, and then resolves the conflicts based on a minimum-distance-to-ideal-point rule using the weighted ranks.

CLUE uses a ranking dynamic modeling approach that locally selects most suitable land use accounting for the aggregate land claims as a driver of the relative competitiveness of the different land use types. Additionally, conversion elasticities, which are established by the user (expert knowledge, only applied if the pixel is already under the specific land use type in the current modeling step) and by a conversion matrix to indicate unlikely conversions as well as fixed conversion trajectories, are accounted for in the model. Local suitability, conversion elasticity and the relative competitiveness represent the land rent for each use at each location, and competitiveness is updated during an iterative procedure to match the allocation and land claims.

DINAMICA normalizes the probability maps of concurrent transitions (the probabilities of transitions concerning the same initial (from) LUC category must sum to one). Next, DINAMICA uses two cellular automata-based transition functions that employ a stochastic selection algorithm: pixels are ranked according to their change potential from greatest to lowest potential. A pruning factor is multiplied by the expected number of cells to be changed and selects the cells that will take part in the selection mechanism based on their spatial probability. Therefore, increasing the pruning factor allows simulated changes to occur in less likely areas. By default, the pruning factor value is ten, which means that ten times the number of cells to be changed are selected based on their transition probabilities and eventually

selected using a lottery process taking into account their change potential and a random number. If the pruning factor is reduced to one, the model becomes deterministic.

### **3.1.5. Reproduction of temporal and spatial patterns**

CA\_MARKOV and DINAMICA use a cellular automata (CA) approach to obtain a proximity effect that makes changes occur in the form of patches to simulate landscape patterns. In CA\_MARKOV, the CA reduces the suitability of land away from existing areas of that type, using by default a 5x5 filter, which can be eventually substituted by a user-defined filter. Users can also control the number of cellular automata iterations and, therefore, the effect of agglomeration around pre-existing or newly created patches. The same CA is applied to all transitions. DINAMICA uses two complementary CA: 1) the Expander and 2) the Patcher. The first process is dedicated only to the expansion or contraction of previous patches of a certain category. The second process is designed to generate new patches through a seeding mechanism. The combination of these two CA presents numerous possibilities with respect to the generation of spatial patterns of change. The user can set parameters to control the mean patch size, the patch size variance, and the isotropy for each transition separately. Increasing the patch size leads to a model with a less fragmented landscape; increasing the patch size variance leads to a more diverse landscape, and setting the isotropy greater than one leads to the creation of more isotropic patches. CLUE and LCM do not apply a CA procedure. CLUE has the option to consider neighborhood interaction that influences the suitability maps through spatial filters (Verburg et al., 2004a).

Additional features aimed at reproducing the spatio-temporal patterns in the different models include the possibility of dealing with transition trajectories, sojourn time, saturation effects, the use of dynamic variables and areas where change is restricted. To address fluctuations in change rates, DINAMICA allows the replacement of the Markov matrix at specific steps of the simulation. Certain transitions are deterministic, such as the transition from secondary forest to mature forest, and depend on the sojourn time of the vegetation succession. Only CLUE allows setting a sojourn time for each transition. In CA\_MARKOV and DINAMICA, a procedure to control the sojourn times using time counters can be easily implemented. Certain



transitions stop when the amount of change has reached a given level. For instance, a deforestation front will move forward, resulting in a certain number of remaining forest fragments. In CA\_MARKOV and DINAMICA, a procedure to control this saturation effect can easily be implemented. However, due to its fixed structure, such implementations are not possible in LCM. In CLUE, the saturation effect can be modeled at the expense of other spatial patterns (neighborhood interactions) because the user is allowed to use only one additional logistic regression by transition.

The use of constraints or incentives in certain areas makes it possible to adjust the change potential to particular spatial policies, such as subsidies of agriculture in certain areas, which cannot be derived from the explanatory variables. LCM and DINAMICA allow the use of constraint or incentive areas at a certain time step of simulation. In CLUE, these areas can both be implemented as a hard constraint for all or specific conversions or by supplying a map with pixel values between zero and one so that these maps are compatible with the regression results and can be used as a soft incentive or constraint. In CA\_MARKOV, the incorporation of constraint or incentive areas can be achieved through the elaboration of suitability maps, taking these areas into account. To simulate the effect taking place at a certain time, a script could be written.

### **3.1.6. Model evaluation**

Generally, the evaluation of LUC prospective maps is based on the comparison between the simulated and the observed (true) map. IDRISI offers various methods to assess the simulations' soft and hard results: 1) the area under ROC (Relative Operating Characteristics) curve, 2) a modified Kappa agreement index (Pontius, 2000) and 3) a validation based on a three-way cross-tabulation. The ROC analysis is a widely used quantitative method to measure the degree to which the presence of a Boolean variable (e.g. the observed change) is associated with high ranks for the change potential variable (Pontius and Parmentier, submitted). The three-way cross-tabulation based validation uses the LUC map of the first date and both, simulated and observed maps (second date) to focus the comparison only on the simulated and observed changes, because simulated and true maps can present a large proportion of

coincidence, due to persistence, although changes are poorly modeled (low coincidence between true and simulated changes). DINAMICA allows computing a fuzzy similarity index based on the concept of fuzziness of location, in which coincidence is not restricted to a strict, cell-by-cell overlay but also includes the cells in a neighborhood (Hagen, 2003). Comparisons can be conducted by applying the fuzziness to the simulated and the true maps of change alternatively. As simulated maps with scattered small patches tend to score higher, the minimum fit value from the two-way comparison is used in order to obtain a conservative assessment of the model. In CLUE, the assessment of the logistic regressions is generally based on ROC analysis, but this assessment has to be carried out using an external program. Applications of CLUE have been assessed using a wide range of methods, such as through the fuzzy equivalent of the Kappa statistic, which evaluates the fuzziness of the location and the category of the simulated land-use patterns (Hagen, 2003) and through a multiple resolution procedure, which evaluates the model performance by quantifying the degree of similarity between the simulated and observed land-use pattern over a range of resolutions (Verburg et al., 2003; Castella and Verburg, 2007). Other methods of assessment, such as the figure of merit (Pontius et al, 2008), can be easily implemented in software packages with GIS operators (DINAMICA and IDRISI).

Mas et al. (2012) pin-pointed the limitations of using only indices based on spatial coincidence when assessing prospective LUC maps. Additional computational criteria, such as landscape indices, can easily be evaluated in programs allowing GIS operations, such as DINAMICA or IDRISI, or these criteria can be carried out using external software packages, such as FRAGSTATS or Patch Analyst (McGarigal et al., 2012; Elkie et al., 1999).

### **3.1.7. Advanced options**

The elaboration of complex models involves splitting the study areas into various subregions that can present different dynamics (e.g., different rates of change, different types of transition, different explanatory variables or/and different effects of the same variables). In CLUE, the user can provide a map of regions that are associated with the corresponding demand for each LUC class and logistic regression models, which define the change potential. DINAMICA

also has the option to divide the study area into regions with particular specifications and parameters. Only DINAMICA is able to run subregion-based models with interactions between the subregions enabling certain variables to have an effect on certain subregions only and variables based on distance to affect the entire study area (e.g., the proximity of a deforestation front in one subregion can influence deforestation in neighboring subregions). In CLUE and IDRISI, subregions can be obtained by running independent models for each and, as a following step, mosaicking the simulated maps, but subregions will not interact, and incompatibilities on the boundaries between regions usually arise.

Dynamic variables are variables that are updated at each time step of the simulation, allowing certain simulated events in previous steps to have an effect on posterior changes. Only CA\_MARKOV does not allow for dynamic variables. Another aspect is the integration of different LUCC patterns and the amount of change over time. This task can be easily performed using DINAMICA, which allows for the substitution of virtually all parameters at given time-steps during simulation, such as the matrices of transition or explanatory variables. For example, Carlson et al. (2012) used different transition matrices for El Niño–Southern Oscillation (ENSO) and non-ENSO years because such years present different patterns and rates of change. IDRISI’s macro modeler can be used to carry out such modeling procedures using CA\_MARKOV. In CLUE, the explanatory variables can be changed in certain time steps, but the regression equations cannot be changed. LCM allows changing some variables during modeling (infrastructure and spatial constraints/incentives), and a substitution of the matrix can be made by concatenating different independent models over time, using the output of one model as the input of the subsequent model. Additionally, DINAMICA allows simulation to be performed in different ways using conditional execution functions (“if then”, “while”).

LCM can use a fixed transition matrix from an exterior model instead of the Markov matrix. DINAMICA can be coupled with an external model that calculates dynamic transition rates and passes them on to the model. This type of external model can be scripted by taking advantage of the operators dealing with images, tables and values, which allows for interaction with the simulated map of previous steps, or running an external process inside DINAMICA. In CA\_MARKOV and LCM, certain transition dynamics could be included by concatenating

models and splitting the simulation horizon into several periods. This concatenating can be performed more easily with CA\_MARKOV because this module can be integrated into a script (macro language) or a macro modeler, which is a graphical modeling interface that enables dynamic linking of GIS modules and raster maps in an algorithmic chain.

Due to the fact that road network is a strong predictor of the spatial patterns of tropical deforestation, but maps of unplanned roads are not available, predictive modeling of the development of roads was developed in DINAMICA and LCM. New road end-points are stochastically selected in areas with highest change potential and are connected to the existing road network using friction maps (e.g. related to topography) in order to achieve the least-cost path and/or to link various areas of high change potential.

LCM provides tools aimed at assessing the impact of change for ecological sustainability and conservation planning, such as tools for species-specific habitat assessment and change studies, gap analysis, landscape pattern evaluation, biodiversity analysis and CO<sub>2</sub> emission assessment. DINAMICA provides some additional tools to model wood harvest volumes processed by sawmills and carbon pool mapping.

Table 2 summarizes the mean characteristics of the methods and tools implemented in each program packages to carry out the main modeling tasks

Table 2 - Methods implemented in each program package.

| Tasks  | CA_MARKO<br>V                                  | CLUE-S                            | DINAMICA                           | LCM   |
|--|--|-----------------------------------|------------------------------------|---|
| <b>Calibration</b>                                   |  |                                   |                                    |   |
| <b>Quantity of change</b>                            |  |                                   |                                    |   |
| Markov   | y  | n                                 | y                                  | y   |
| Alternative approaches easily implemented?           | n  | y                                 | y                                  | n   |
| <b>Change potential evaluation</b>                   |  |                                   |                                    |   |
| Data driven statistical approach                     | -  | Log Reg                           | W of E                             | MLP, SimWeight  |
| Data driven machine learning approach                | -  |                                   | Gen Alg                            | Log Reg   |
| Knowledge driven approach                            | MCE  |                                   | Weight edition                     |   |
| <b>Simulation</b>                                    |  |                                   |                                    |   |
| <b>Spatial allocation of change</b>                  |  |                                   |                                    |   |
|  | multi-objective allocation                     | ranking dynamic modeling approach | stochastic selection algorithm     | multi-objective allocation                              |
| <b>Reproduction of temporal and spatial patterns</b> |  |                                   |                                    |   |
| Landscape patterns simulation                        | CA easily implemented through suitability maps | Optional spatial filters          | CA easily implemented              | -   |
| Sojourn time   |  | yes                               |                                    | no  |
| constraints or incentives                            |  | yes                               | yes                                | yes   |
| <b>Assessment</b>                                    |  |                                   |                                    |   |
| <b>Model evaluation</b>                              |  |                                   |                                    |   |
| Alternative approaches easily implemented?           | ROC, Kappa                                     | ROC (outside CLUE)                | Fuzzy similarity index             | ROC, Kappa  |
|  | yes  | outside CLUE                      | yes                                | yes   |
| <b>Advanced options</b>                              |  |                                   |                                    |   |
| Subregion  | running independent models                     | running independent models        | yes                                | no  |
| Dynamic variables                                    | no   | yes                               | yes                                | yes   |
| Dynamic change rates                                 | concatenating models                           | from external model               | from external or internal submodel | from external transition matrix or concatenating models |
| Road modeling  | no   | no                                | yes                                | yes   |

### **3.1.8. Additional considerations**

As additional considerations, we assess the flexibility, user friendliness, and cost.

Flexibility involves the possibility of adapting the model to the user's needs and the possibilities of interaction with other programs. The programs offering more flexibility in the development of the modeling step of model fitting according to user needs are CA\_MARKOV and DINAMICA. Models can be constructed taking advantage of the large pool of tools and operators available in the program environments. Moreover, programming is easy even for users without previous programming experience because a graphical interface is provided in which operators can be dragged and linked dynamically to integrate feedback and iterative operations. In DINAMICA, advanced modelers can directly develop their models using the EGO or XML programming languages. LCM and CLUE present a rigid structure that defines a fixed flow of procedures with respect to LUCC modeling. This structure can be advantageous for new users because the procedures are well defined and documented. However, it becomes a drawback when users want to develop "custom-made" or more complex models. In general, all of the models, with the exception of CA\_MARKOV, can make use of external data, and all can be validated using different freely chosen methods that are independent from the integrated assessment module. DINAMICA and IDRISI offer the possibility of running external processes.

User friendliness refers to the features that make the use of the model easier, such as employing a particular tool and offering explanations in the documentation provided to the user. This characteristic is rather subjective, as it also depends on the user's previous experience, background and preference. The IDRISI programs (CA\_MARKOV and LCM) maybe more user-friendly than the others because 1) they are well documented (IDRISI manual and tutorial, IDRISI discussion forum, tutorial videos), 2) all of the operations can be executed in a window interface environment but can also be automated through the script and programming tools and 3) operations prior to modeling (e.g., image classification to create LUC maps) can be performed in the same environment. DINAMICA also has an intuitive programming interface and is very well documented (user guide, video guides, wiki page and

discussion list). Recent versions include the possibility of using or writing a wizard, which a sequence of pages that may contain text, tables, images, diagrams and videos (similar to a PowerPoint presentation). Each page can contain general information about the model and the simulation inputs that are directly linked to the script inputs, which enables the user to set parameters and run simulations. CLUE is also relatively well documented (user guide with exercises and a large amount of scientific publications) but requires some laborious manual operations (e.g., text file editing to enter parameters) and carrying out operations using external programs (logistic regression, map displaying, model input and assessment). This lack of integration and visualization also makes it very difficult to find errors in the model implementation. DINAMICA and IDRISI allow importing and exporting a large amount of raster formats, while CLUE only accepts ASCII files. A DINAMICA operator allows reading and writing directly from/to zip files as well as defining a root folder, which makes it easier to share and transport models across computers with different folder structures. It is worth noting that user friendliness can be a counterproductive factor because the possibility of elaborating a LUCC model easily and rapidly can lead to poorly informed applications.

All of the software programs assessed are designed for Microsoft Windows and run on 32- and 64-bit platforms. IDRISI is a hybrid 32/64-bit program, and CLUE requires a 32-bit system. DINAMICA can be installed as a native 64-bit program, which allows the use of the expanded memory address and performance optimizations. This program is able to handle multiple sets of large maps (up to 64,000 x 64,000 cells) because of its mechanism of disk paging. It enables multiprocessor architecture to split execution pipelines on different processors and run several operators' internal algorithms using parallel processing; models can also be executed from a command line, which increases the model speed.

CLUE and DINAMICA are available for use at no cost. IDRISI, is a relatively low-cost software package and presents a large number of GIS and image processing modules which can be used in other tasks.

### **3.2. The virtual case study**

The virtual case study was inspired by real LUCC situations (Mas and Flamenco, 2011; Pérez Vega et al., 2012) and included two LUC maps (dated 2000 ( $t_0$ ) and 2010 ( $t_1$ ), 200 x 300 pixels) and only a few explanatory maps (elevation, distance to roads, distances to previous LUC) available at <http://www.ciga.unam.mx/>. Four different LUC categories were distinguished for the simulation: forest (F), two types of cropland (CL1 and CL2) and urban area (U) (Figure 3). During 2000-2010 (the calibration period), only four types of transitions were produced, including two different deforestation patterns (forest to croplands), forest regeneration, which followed the abandonment of cropland 1, and urban growth. Patches of cropland were created through both the expansion of previous patches and the creation of new patches. Cropland 1 was associated with low elevations ( $< 600$  m) and a close proximity to roads and urban area, although some conversions of forest to cropland 1 were located in remote areas. An unexpected, large amount of change from forest to cropland 1 was placed in a particular distance range from roads (1650-1700m). Cropland 2 was limited to an optimal elevation range between 850 and 1100 m. Urban areas were expanded only at the expense of cropland 1 surrounding the existing urban area. The four software packages were applied to the virtual LUCC case study using a basic past trend-based simulation to create prospective LUC maps for 2020 ( $t_3$ ) to illustrate their different tools and settings.



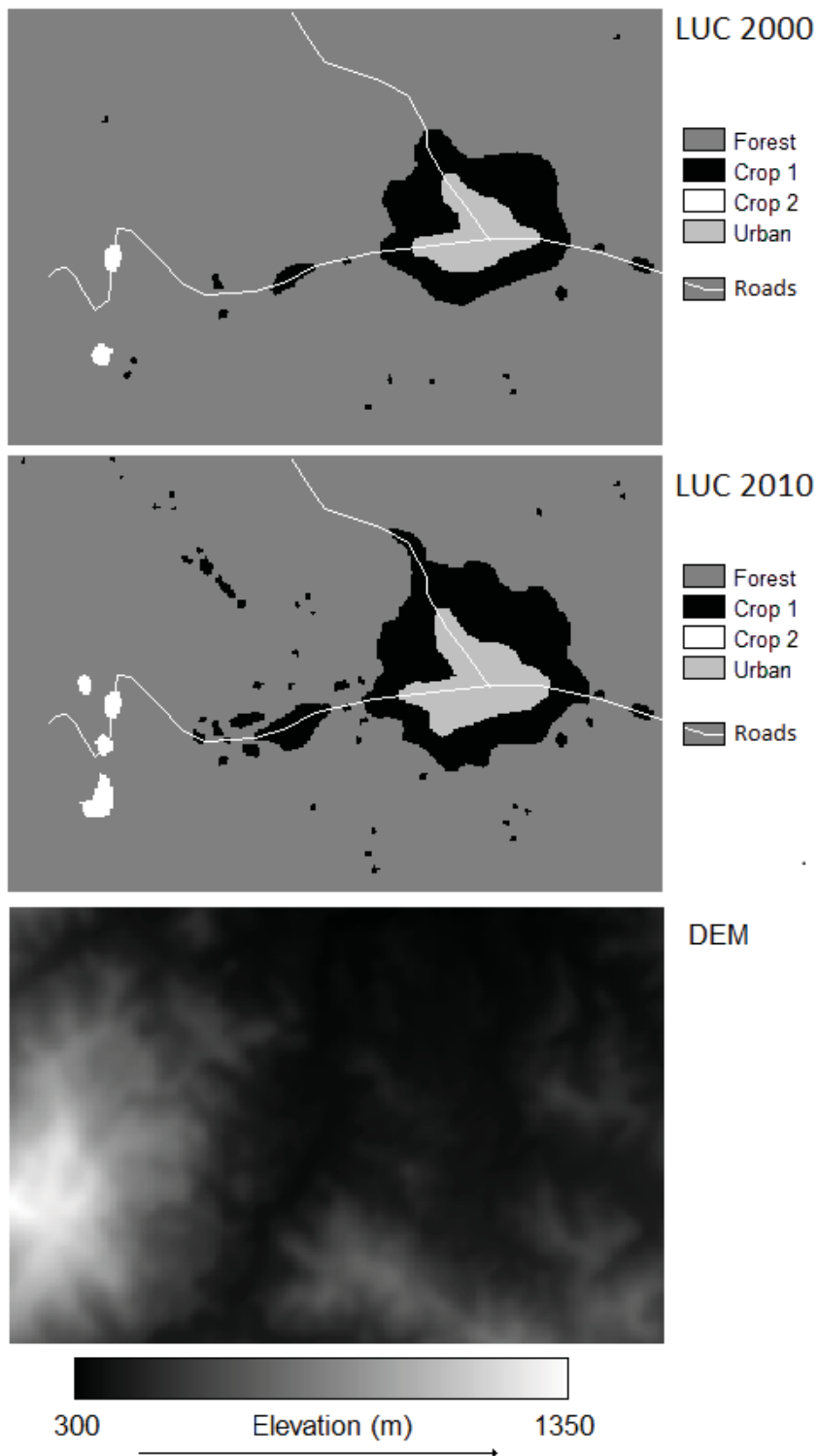


Figure 3 – Virtual LUCC case. The LUC maps were created following the rules about LUC distribution and LUCC patterns described in section 3.2

### 3.2.1. Quantity of change estimate

By overlaying the two LUC maps from 2000 and 2010, a transition matrix was generated that depicts four change transitions: 1) forest to cropland 1, 2) forest to cropland 2, 3) cropland 2 to forest and 4) cropland 1 to urban (table 3). Tables 4 and 5 show the base transition matrix (for the entire 10-year period) and the annual matrix, respectively, and was obtained from DINAMICA using equation (1).

Table 3 – Transition matrix 2000-2010 in number of pixels (change pixels are in bold)

| 2010<br>2000 | F         | CL1         | CL2        | U          | Total |
|--------------|-----------|-------------|------------|------------|-------|
| F            | 51799     | <b>2725</b> | <b>201</b> | 0          | 54725 |
| CL1          | <b>33</b> | 3190        | 0          | <b>602</b> | 3825  |
| CL2          | 0         | 0           | 149        | 0          | 149   |
| U            | 0         | 0           | 0          | 1301       | 1301  |
| Total        | 51832     | 5915        | 350        | 1903       | 60000 |

Table 4 – Base transition matrix (entire period) obtained by DINAMICA

| 2010<br>2000 | F      | CL1    | CL2    | U      |
|--------------|--------|--------|--------|--------|
| F            | 0.9465 | 0.0498 | 0.0037 | 0      |
| CL1          | 0.0086 | 0.8340 | 0      | 0.1574 |
| CL2          | 0      | 0      | 1      | 0      |
| U            | 0      | 0      | 0      | 1      |

Table 5 – Annual transition matrix obtained by DINAMICA

|     | F        | CL1      | CL2      | U        |
|-----|----------|----------|----------|----------|
| F   | 0.994088 | 0.005536 | 0.000376 | 0        |
| CL1 | 0.000959 | 0.981983 | 0        | 0.017058 |
| CL2 | 0        | 0        | 1        | 0        |
| U   | 0        | 0        | 0        | 1        |

The projected areas for 2011 to 2030 were obtained by iteratively applying the annual matrix to the areas of the previous year. A set of Markov probability matrices was calculated using IDRISI to calculate areas for each projected year during the same period. As IDRISI provides a method of correction to compensate for the classification errors (see section 3.1.2. Quantity of change estimate), two sets of matrices were created (without correction and with correction to 15% error maps, Tables 6 and 7).

Table 6 – Transition matrix generated by IDRISI to project one year after the end of the calibration period (2011). Note that due to the IDRISI method, this matrix is not equivalent to the annual matrix of DINAMICA. It allows estimating LUC areas of 2011 based upon 2010 but it cannot be iteratively used 10 times to obtain 2020 LUC area estimates, for example. Unobserved transitions (value of zero) are indicated in bold characters.

|     | F        | CL1      | CL2      | U                 |
|-----|----------|----------|----------|-------------------|
| F   | 0.994300 | 0.005400 | 0.000300 | 0                 |
| CL1 | 0.001200 | 0.978100 |          | <b>0</b> 0.020700 |
| CL2 |          | <b>0</b> | <b>0</b> | 1 <b>0</b>        |
| U   |          | <b>0</b> | <b>0</b> | <b>0</b> 1        |

Table 7 – Transition matrix generated by IDRISI to project one year after the end of the calibration period (2011), including the option to correct estimate considering the classification error of 15% in the input map. Note that some unobserved transitions exhibit high probabilities (in bold characters).

|     | F               | CL1             | CL2             | U               |
|-----|-----------------|-----------------|-----------------|-----------------|
| F   | 0.845100        | 0.146900        | 0.007900        | 0.000000        |
| CL1 | 0.009500        | 0.831400        | 0.000000        | 0.159100        |
| CL2 | <b>0.050000</b> | <b>0.050000</b> | 0.850000        | <b>0.050000</b> |
| U   | <b>0.050000</b> | <b>0.050000</b> | <b>0.050000</b> | 0.850000        |

A projection of areas of LUC for a medium term (20 years, see figure 4) does not show important differences between DINAMICA and IDRISI (option without using the error correction). When LUC areas are projected to 2030, the differences between the projected areas in both approaches ranges between 0.2% (Cropland 1) and 10.0 % (Urban). In practice,

the differences between IDRISI (without correction) and DINAMICA do not affect the accuracy of the estimation of change, especially when considering that the assumption of the linearity of LUCC is generally wrong (e.g., the assumption that trends of change during the calibration period are similar to the following period of time). When the correction used by IDRISI to correct errors in the maps is applied, this correction affects the projected area estimations significantly (Figure 4). Differences in the projected areas for 2030 calculated by IDRISI with and without correction ranging from 14.6% (Forest) to 73.1% (Cropland 2) are observed. The difference in the projected areas between DINAMICA and IDRISI (without correction) increases slowly over time because the same annual matrix is applied iteratively (that is, areas for 2030 were obtained applying this matrix first to the 2010 areas, a second time to the 2011 projected areas, and so on). The difference between these two projections and the projection obtained by IDRISI with correction is stable over time because the projected areas are obtained by applying different matrices to the same vector of areas at the end of the calibration period (the areas of 2010, in our case).

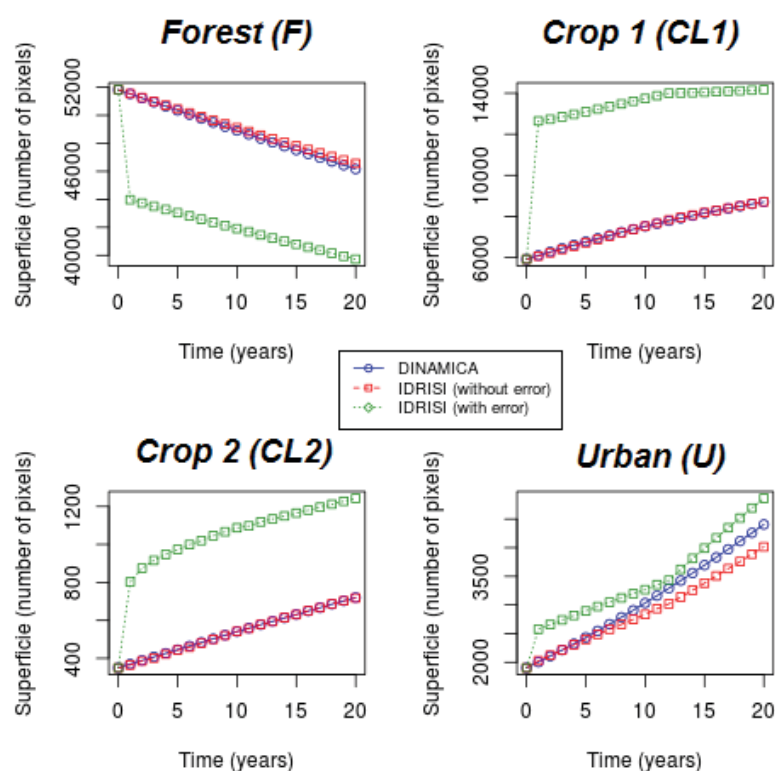


Figure 4 - Projected areas with DINAMICA and IDRISI (both with and without correction of LUC maps error).

In some cases, equation 1, used by DINAMICA to obtain the annual transition matrix, does not result in a solution that is expressed with real numbers and DINAMICA is not able to provide such a matrix (for details see Takada et al., 2010). The advantage of the IDRISI approach resides in the fact that it does not use matricial calculation and is always able to provide a result. Conversely, the correction performed by IDRISI, to correct the bias related to the mapping errors, modifies all the transition probabilities with probability  $> 0$  (and probabilities of 0 if only the permanence transition has probability  $> 0$ , table 7), and the user has to be sure that these values are realistic. In this case, tables 6 and 7 show that, after this correction, all permanence probabilities are lowered, and impossible or unobserved transitions present high probabilities (for instance, the transitions urban and CL2 to other categories exhibits a probability of 5%, table 7).

### **3.2.2. Change potential evaluation**

Maps of the change potential were constructed using weight of evidence and genetic algorithm (DINAMICA) as well as a Multilayer Perceptron (MLP), SimWeight and logistic regression (LCM). Neural network training was carried out with the default setting (learning rate from 0.01 to 0.001, momentum 0.5, number of hidden nodes calculated as the average between numbers of input and output nodes, 10,000 iterations). In the case of overfitting, the number of iterations was reduced. In the elaboration of the logistic regression models, stepwise selection of variables was performed outside of IDRISI using the Akaike's Information Criterion (AIC), a measure of the relative quality of a statistical model that enables to manage a tradeoff between goodness of fit and model complexity and expert knowledge about the importance of each variable on the distribution of the LUCC.

For CA\_MARKOV and CLUE, suitability maps for the two types of cropland, forest and urban areas were elaborated using multi-criteria evaluation (MCE) and the logistic regression method, respectively. For CA\_MARKOV, suitability maps were constructed using linear and sigmoid functions to establish a relationship between suitability and explanatory variables. Then, suitability maps for the same LUC were combined by weighing them using the

analytical hierarchy method (a pairwise comparison technique, see Saaty, 1997) to form a single, final suitability map using MCE. The weighting process allowed an evaluation of tradeoffs so that more weight was given to distance based factors, like distance to road or distance to a specific LUC, as to DEM or DEM derived factors, such as slope (both less than 10 %). For CLUE, the logistic regression models were developed using the same criteria used for the change potential maps. We used the same variables in the models based on change potential maps except when other explanatory variables were more appropriate to the models according to AIC index. As shown in table 8, the different approaches can lead to selecting different explanatory variables to elaborate the change potential maps.

Table 8 – Variables used to elaborate the change potential and suitability maps

|                                     | Transition / category | Elevation | Distance to roads | Distance to forest | Distance to Cropland1 | Distance to Urban | Distance to No Forest |
|-------------------------------------|-----------------------|-----------|-------------------|--------------------|-----------------------|-------------------|-----------------------|
| Change Potential (DINAMICA and LCM) | F to CL1              | x         | x                 |                    | x                     |                   |                       |
|                                     | F to CL2              | x         |                   |                    |                       |                   |                       |
|                                     | CL1 to F              |           | x                 |                    |                       | x                 |                       |
|                                     | CL1 to U              |           | x                 |                    |                       | x                 |                       |
| Suitability (CA_MARKOV and CLUE)    | F MCE                 |           |                   |                    | x                     | x                 | x                     |
|                                     | F CLUE                |           | x                 |                    | x                     |                   |                       |
|                                     | CL1 (MCE and CLUE)    | x         |                   |                    |                       | x                 |                       |
|                                     | CL2 (MCE and CLUE)    | x         |                   |                    |                       |                   |                       |
|                                     | U (MCE)               |           |                   |                    |                       | x                 |                       |
|                                     | U (CLUE)              |           | x                 | x                  |                       |                   |                       |

Figure 5 shows the relationship between two explanatory variables and the change potential maps for the transition “Forest to cropland 1” or suitability values for Cropland 1. Figure 5(f) shows the proportion of pixels of forest that experienced the transition. It is important to note

that the two suitability maps (Figure 5 (g) and (h)) were elaborated using, as explanatory variables, “Elevation” and “Distance to urban”, and in contrast, all the change potential maps (Figure 5 (a) to (e)) were based upon the explanatory variables “Elevation”, “Distance to Cropland 1” and “Distance to Roads”. It can be observed that the relationship between the change potential and the explanatory variables obtained by the weights of evidence presents a more complex (irregular) shape than the other approaches. An intermediate complexity is given by SimWeight and very similar, almost monotonic shapes are obtained by logistic regression and the MLP. Figure 6 indicates that almost all the maps are highly correlated. The change potential maps obtained by LCM using MLP and logistic regression are very similar (Spearman = 0.92), and the main difference is that the values of the high change potential of the MLP tend to saturate. The higher change probabilities ranging from 0.3 to 0.5 in the logistic regression correspond to one probability in the MLP change potential map. The methods of weights of evidence (with or without GA adjustment) and the MLP generated highly correlated maps (Spearman = 0.79). The two suitability maps are also highly correlated (Spearman = 0.72) and present low correlation with the change potential maps, mainly because they are based on different explanatory variables.

In this exercise, the weights of evidence were not edited, and a visual editing consisting of smoothing the relationship between the weights and the explanatory variables (eliminating noisy fluctuations of weight values) is expected to lead to more regular shapes. Conversely, a more complex MLP (more nodes in hidden layers) could lead to a more complex shape. The ability to use complex functions to establish the relationship between change and explanatory variables is expected to control the goodness of fit of such a relationship in addition to the tendency to overfit.

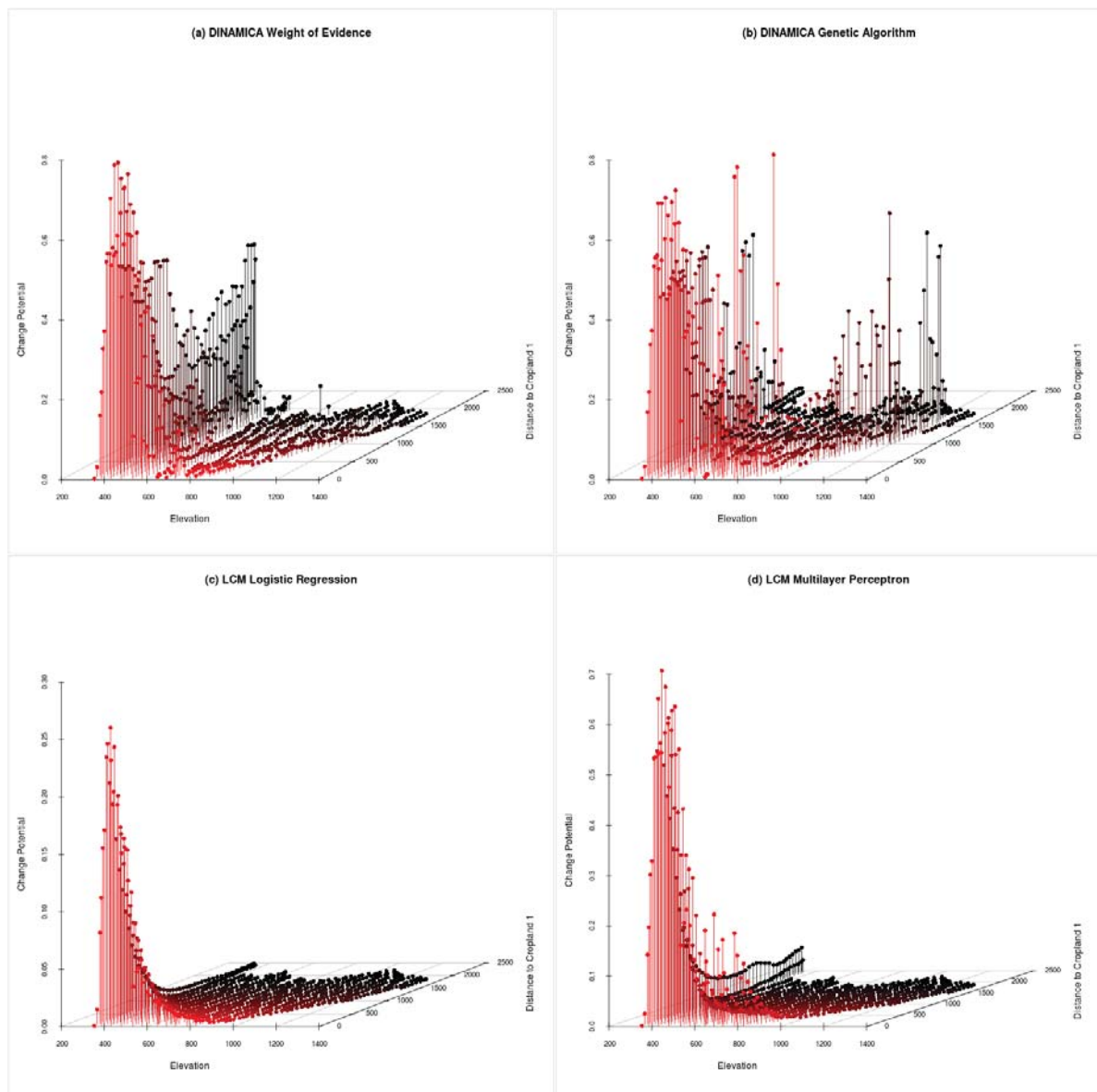


Figure 5 (part 1) – Scatterplot representing the relationship between change potential (or suitability) and two explanatory variables with the different calibration methods (transition “Forest to cropland 1” and cropland 1).



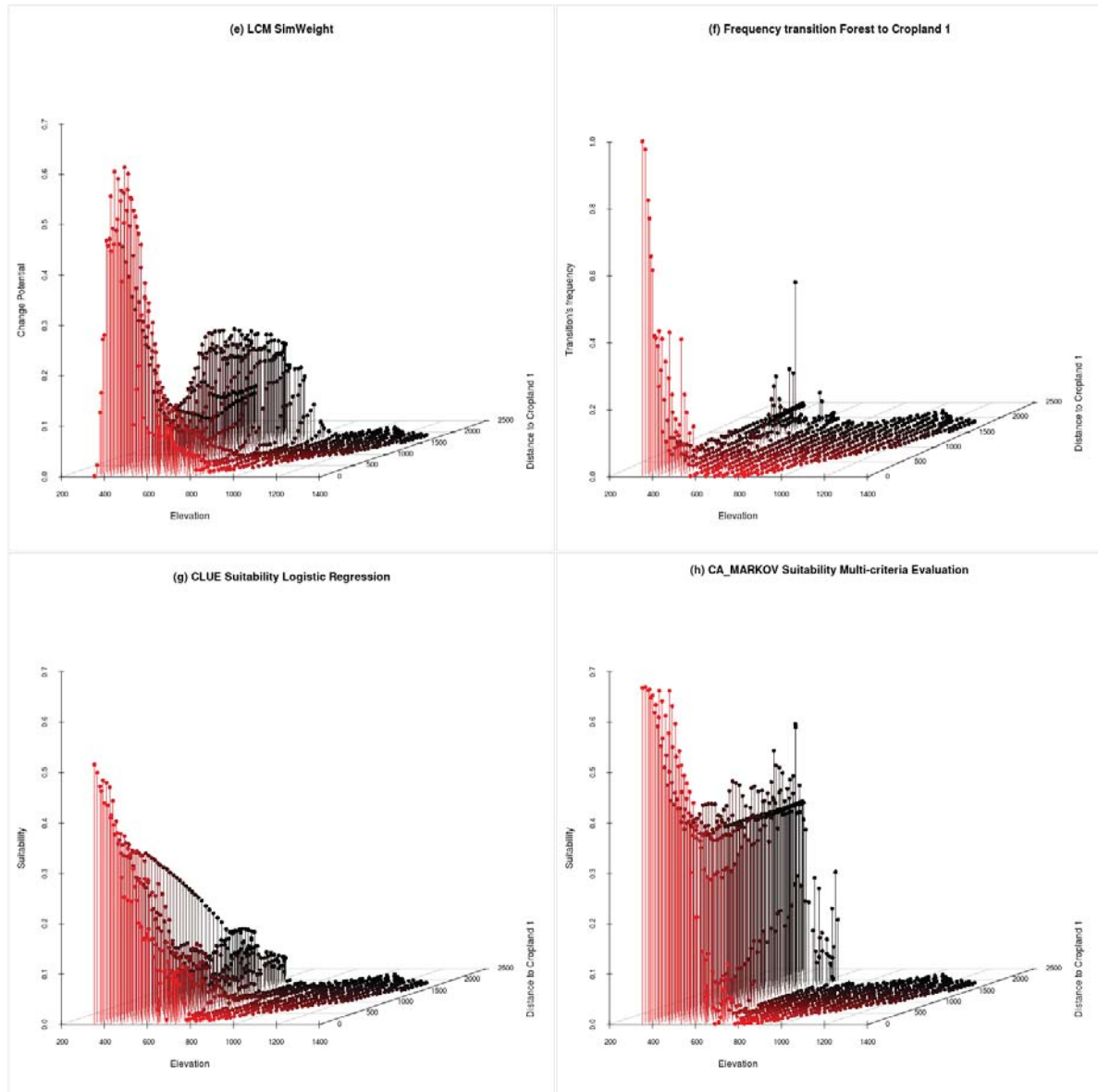


Figure 5 (part 2) – Scatterplot representing the relationship between change potential (or suitability) and two explanatory variables with the different calibration methods (transition “Forest to cropland 1” and cropland 1).

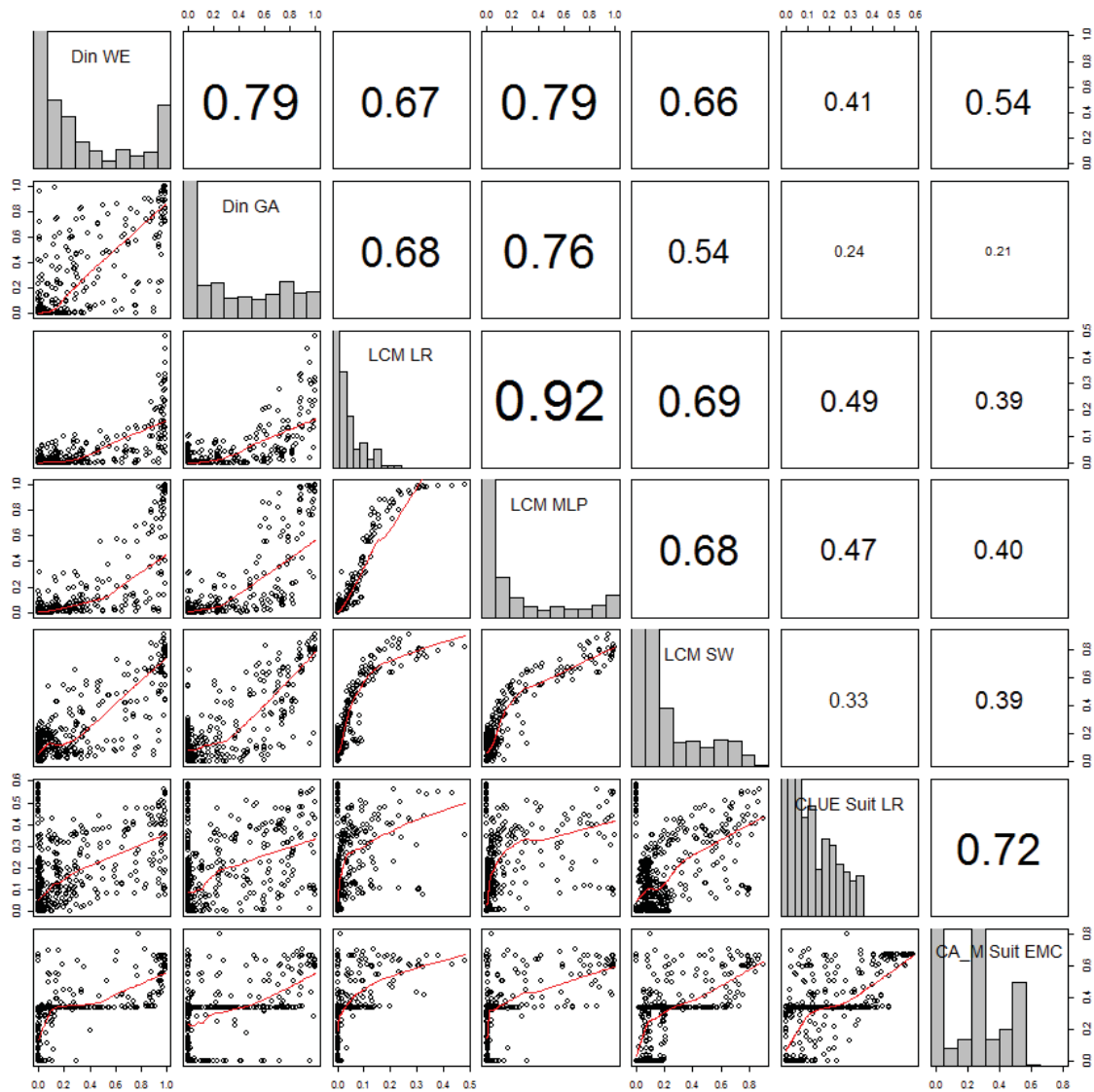


Figure 6 - Change potential of the transition “Forest to cropland 1” and cropland 1 suitability (CLUE, CA\_MARKOV). The lower part of the figure shows a scatterplot of the values of change potential or suitability of the 7 approaches examined. Histograms of the frequency of the values obtained by each approach are presented in the diagonal. The upper part of the figure indicates the Spearman coefficient of correlation between the values of pair of approaches.

To assess these two last aspects, we conducted two experiments. As a first step, In order to test the likely tendency of the different approaches to overfit the probabilities, the change potential of forest to cropland 1 was examined as a function of the distance from roads (Figure 7). Only

the weights of evidence (both weights of evidence only and the weights adjusted by genetic algorithm) show a clear increase in the change potential values for the range of distance where a large amount of this transition was placed. This can be attributed to the fact that the weight of evidence for a particular explanatory category is calculated taking into account the frequency of change and no change inside or outside this category. The other methods try to adjust a function by taking into account the values and the observed changes of all the categories. As a consequence, the weight of evidence method is prone to overfit more easily than methods known to be highly flexible, such as neural networks. In the case of DINAMICA, a visual inspection of the relationship between weight values and explanatory variable categories can be easily performed in the weights display tool to detect and correct such cases of overfitting.

To detect the opposite effect, specifically an overgeneralization of the relationship between the explanatory variables and the change potential/suitability, the change potential/suitability of the transition “Forest to cropland 2”/category crop land 2 was graphed as a function of the elevation (Figure 8). It can be observed that only the weights of evidence (DINAMICA), the SimWeights method (LCM) and the multicriteria evaluation (CA\_MARKOV) were able to reproduce the optimal range of elevation for cropland 2. The sigmoid curve of the logistic regression was not able to replicate this pattern. The MLP used to produce the change potential map in LCM had only one hidden layer with two nodes, and allowed for the correct classification of more than 90% of the training data but was unable to replicate the elevation pattern. An MLP with more nodes or hidden layers would likely replicate the elevation pattern but would be more likely to overfit.

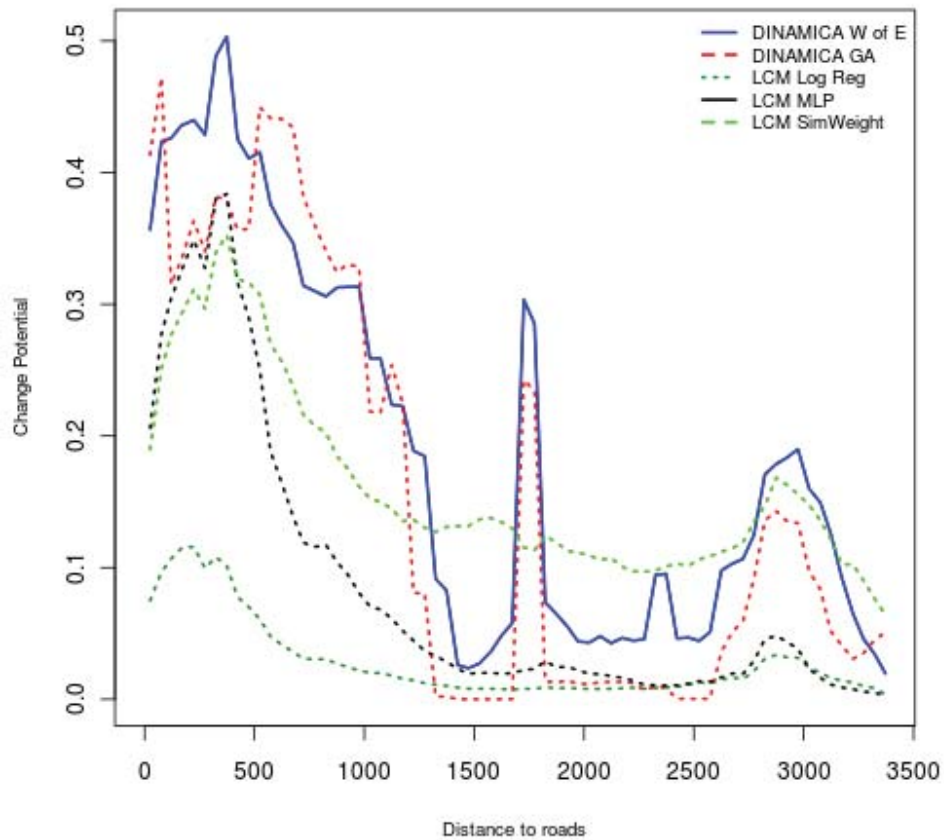


Figure 7 – Relationship between change potential of transition “Forest to cropland 1”/category Cropland 1 and distance to road. DINAMICA change potential maps exhibit an important increase of their values around the distance 1650-1700 m where a large area of "Forest to cropland 1" was placed.

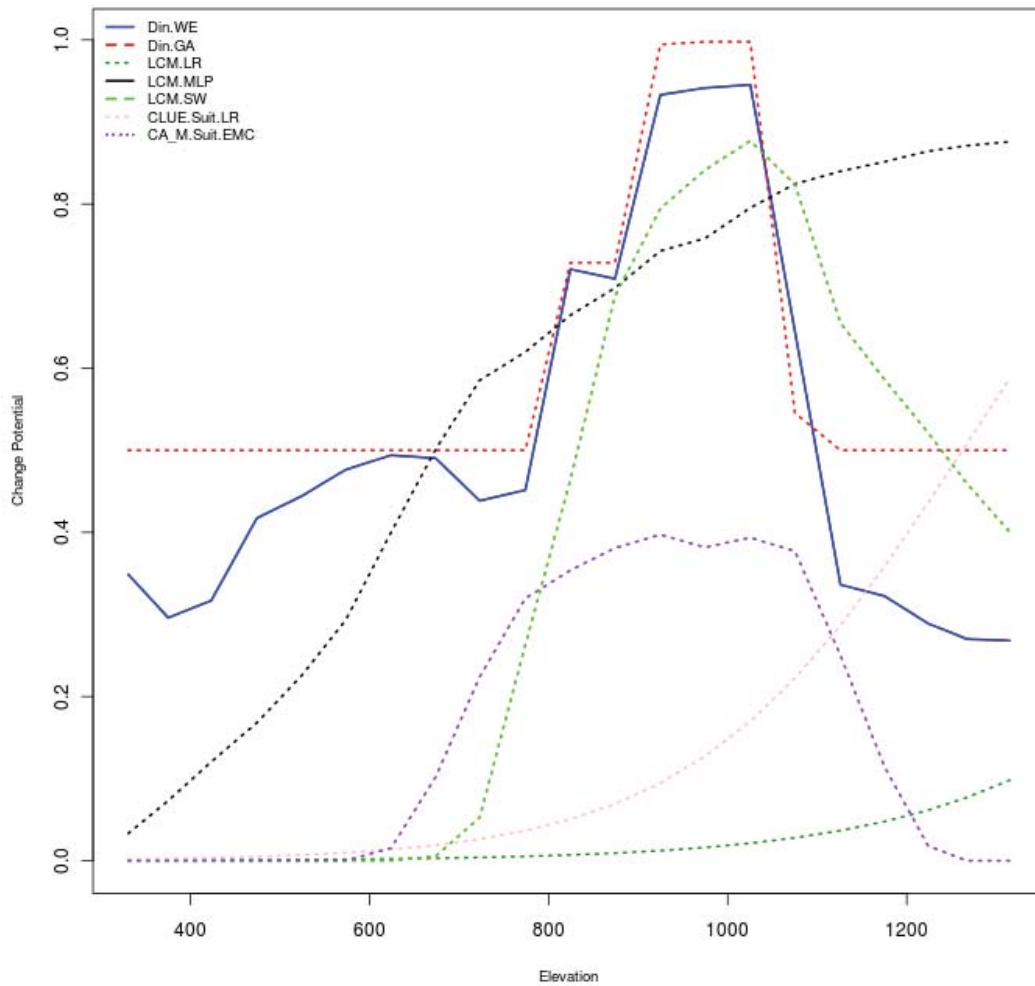


Figure 8 – Relationship between change potential/suitability of transition “Forest to cropland 2”/category Cropland 2 and elevation. Only DINAMICA and LCM using SimWeight were able to shape the optimum range of elevation between 850 and 1100 m approximately.

### 3.2.3. Allocation of change and reproduction of temporal and spatial patterns

Prospective maps for 2020 were produced using the different software packages. The same quantity of change (derived from DINAMICA matricial computing) was used in order to make the comparison easier (Figure 9). Although the general pattern of the different maps is similar because all of the simulations start from the LUC map of 2010, they present important

differences. The location of the change is different because it depends on the different change potential or suitability maps. For instance, LCM presents different results with the change potential maps obtained using logistic regression, MLP or SimWeight. In some cases, changes are allocated in unlikely places due to inaccuracies in the change potential maps, as seen, for example, in the patches of Cropland 2 at elevations above 1200 m in LCM with MLP.

Programs without cellular automata were able to create new patches (for example LCM with MLP and CLUE) due to the spatial autocorrelation of the explanatory variables that leads to the formation of small areas of high change potential in the change potential/suitability maps, which are allocated to be changed by the model. Nevertheless, the user cannot control the size and the shape of these patches because they depend only on the spatial distribution of areas with a high change potential and the quantity of change to be modeled. Only DINAMICA can influence landscape pattern simulation using its cellular automata without modifying input data. However, the control of patch characteristics is not an easy task because the landscape structure of the simulated maps depends on the change potential maps and the cellular automata setting (mean patch size, size variance and isometry). It is also worth noting that in DINAMICA, the change allocation is not completely a deterministic process (unless the pruning factor is set to one; see the section on spatial allocation of change), and consequently, each run results in a slightly different result.